

KATALYSIA RESEARCH · PUBLIC



Human-AI-Training und Integration von Spezialwissen in LLMs

Agent-Prompt v1.0 -- Auftragsdefinition fuer die KATALYSIA-Studie

*Verbindlicher Studien-Auftrag mit modularer Fachgebietsstruktur und Querpruefungs-
Checkliste*

Stand: 12. Mai 2026

Basis: Claude Opus 4.6 · KATALYSIA Research

Vertraulich · Interne Arbeitsunterlage · Weitergabe nur nach Rücksprache

AGENT-PROMPT v1.0

Studie: Human-AI-Training und Integration von Spezialwissen in Large Language Models

Themen-Slug: human-ai-training **Stand des Prompts:** 12. Mai 2026 · Version 1.0
Modellbasis (Soll-Erstellung): Claude Opus 4.6 oder neuer **Auftraggeber:** KATALYSIA Research Public **Sprache:** Professionelles wissenschaftliches Deutsch **Verbindliche Designgrundlage:** [KATALYSIA-Designsystem_Verbindlich.md](#) **Master-Referenzen für Layout, Tonalität und Tiefe:** public/data-center/, public/private-wealth-nrw/, public/corporate-finance-advisory/ (insbesondere die Vollfassung v2.0 vom 8. Mai 2026)

1. Auftrag in einem Satz

Erstelle eine umfassende, evidenzbasierte und wissenschaftliche KATALYSIA-Research-Public-Studie zum Thema *Human-AI-Training und Integration von Spezialwissen in Large Language Models* mit den drei Pflicht-Tiers (Ausführliche Studie, Executive Summary, Executive Presentation), die in Inhalt, Layout und Tonalität dem Master public/corporate-finance-advisory/ ebenbürtig ist und so strukturiert ist, dass weitere Fachgebiete (Kapitel 6) später modular ergänzt werden können.

2. Verbindliche Vorgaben (nicht verhandelbar)

- 1. Designsystem.** Vollständige Übernahme der Vorgaben aus KATALYSIA-Designsystem_Verbindlich.md (Designtokens, Cover- und Inhaltslayouts, Pflicht-Schluss-Sektionen, Standard-Textbausteine, Quellenstruktur, Querprüfungs-Checkliste).
- 2. Master-Querprüfung.** Vor der Abnahme jeder Datei Quervergleich gegen public/corporate-finance-advisory/2026-05-08_studie-vollfassung-v2-0.html, 2026-05-08_executive-summary-v2-0.html und 2026-05-08_executive-presentation-v2-0.pptx. Strukturtiefe, Tabellengestaltung, Print-CSS und Cover-Logik sind 1:1 zu übernehmen.
- 3. Ablage.** Alle Deliverables in public/human-ai-training/. Datei-Namensschema: 2026-05-12_<deliverable-slug>-v1-0.<ext>. Eine index.html als Themen-Indexseite mit drei .content-tier-Blöcken.

4. **Stylesheet und Logo.** Ausschließlich `public/assets/style.css` und `public/assets/logo_katalysia_research.png`. Kein eigenes CSS, kein neues Logo.
 5. **Sprache und Stil.** Wissenschaftliches Deutsch, evidenzbasiert, mit transparenten Annahmen, Drei-Szenarien-Bandbreiten (Konservativ / Base / Ambitioniert), keine Marketingsprache. Internationale Anbieternamen und Fachbegriffe (z. B. RLHF, DPO, SFT, Constitutional AI, Frontier Lab) im Original.
 6. **Standard-Textbausteine.** Methodik & KI-Einsatz, Disclaimer, Vertraulichkeits- und Schlussvermerk wörtlich gemäß Designsystem Abschnitt 5 in PPTX, DOCX und HTML.
 7. **Quellen.** Pflicht-Zwei-Spalten-Struktur (Primärquellen — Branchenreports & Industrie / Regulierung, Recht & Statistik) in PPTX-Schlussfolie, DOCX-Schlussseite und HTML-Schluss-Tabelle.
 8. **Drei Pflicht-Tiers.** Ausführliche Studie (Vollfassung; optional zusätzlich Mittel- und Kompaktfassung in .md analog Corporate Finance Advisory), Executive Summary, Executive Presentation — alle Tiers inhaltlich konsistent (Zahlen, Aussagen, Annahmen, Modellversion, Datum).
 9. **Modularität (Kapitel 6).** Das Kapitel zur fachgebietsspezifischen Vorbereitung ist strikt modular aufgebaut: ein normiertes Unterkapitel-Schema, das pro Fachgebiet identisch ist, sodass künftige Fachgebiete (z. B. „Strategy Consulting“, „Investment Banking“, „Compliance“, „Risikomanagement“, „Medizin“, „Pharma & Life Sciences“, „Engineering“) später ohne Re-Architektur hinzugefügt werden können.
-

3. Studienarchitektur (Pflicht-Gliederung Ausführliche Studie)

Die Vollfassung enthält die folgenden Kapitel (zwingend in dieser Reihenfolge und Numerierung). Tiefen-Richtwert: ca. 80–140 Druckseiten Vollfassung (analog Corporate Finance Advisory), Executive Summary 8–12 Seiten, Präsentation 14–18 Folien.

| Kapitel 1 — Einleitung und Forschungsdesign

1.1 Motivation und Erkenntnisinteresse 1.2 Begriffsabgrenzungen (Pre-Training, Mid-Training, Post-Training, Alignment-Tuning, Fine-Tuning, Domain-Adaptation, Knowledge-Distillation, Synthetic-Data-Generation, Inference-Time-Steering) 1.3 Forschungsfragen (FF1 Konzepte und Methoden, FF2 Anwendung in Frontier Labs, FF3 Qualitäts- und Sicherheitswirkung, FF4 Anbietermarkt und Geschäftsmodelle, FF5 Marktgröße und Wachstumsdynamik, FF6 Fachgebietsspezifische Vorbereitung) 1.4 Methodik im Überblick (Desk Research, Vendor-Disclosures, Modellkarten, akademische Literatur, Anbieterwebseiten, Stellenanzeigen-Analyse, Pricing-Beobachtung) 1.5 Untersuchungsrahmen, Zeitraum, Region (Global, Schwerpunkt USA / DACH / EU), Limitationen

Kapitel 2 — Konzepte und Methoden des Human-AI-Trainings

2.1 Lebenszyklus eines LLM: Pre-Training → Mid-Training → Post-Training → Continual-Learning 2.2 Supervised Fine-Tuning (SFT) und Instruction-Tuning — Datenformate, Qualitätskriterien, typische Volumina 2.3 Präferenzbasiertes Lernen 2.3.1 Reinforcement Learning from Human Feedback (RLHF) — Pipeline (SFT → Reward-Model → PPO/GRPO) 2.3.2 Direct Preference Optimization (DPO) und Varianten (IPO, KTO, ORPO, SimPO) 2.3.3 Reinforcement Learning from AI Feedback (RLAIF) und hybride Schemata 2.4 Constitutional AI und prinzipiengeleitetes Alignment (Anthropic-Schule) 2.5 Reasoning- und Tool-Use-Training — verifikable Rewards (RLVR), Process Reward Models, Self-Play, „Verifizier-Generator“-Architekturen 2.6 Domain-Adaptation und Spezialwissen-Integration 2.6.1 Continued Pre-Training auf Fachkorpora 2.6.2 Expert-Authored Datasets (Aufsätze, Lösungswege, Code-Reviews, Bewertungen) 2.6.3 Synthetic Data Pipelines unter Experten-Validierung 2.6.4 Retrieval-Augmented Generation (RAG) und tool-gestützte Lösungen als Komplement 2.7 Sicherheits- und Red-Team-Training — Adversarial Probing, Refusal-Training, Jailbreak-Resistenz 2.8 Evaluation und Benchmarks — MMLU, GPQA, MATH, HLE, SWE-Bench, ARC-AGI, fachspezifische Benchmarks (CFA, MBE/Bar, Steuerberater-Examen, CPA), Expert Win-Rates, A/B-Tests 2.9 Qualitätskriterien für Trainingsdaten — Korrektheit, Vollständigkeit, Begründungstiefe, Stil, Diversität, Bias-Kontrolle

Kapitel 3 — Trainingsstrategien führender LLM-Anbieter

Je Anbieter: Trainingsphilosophie, öffentlich dokumentierte Pipeline, Human-Feedback-Programme, bekannte Datenpartner, Differenzierungsmerkmale, Sicherheits-/Alignment-Schule, öffentliche Performance-Belege. 3.1 Anthropic (Claude Opus / Sonnet / Haiku; Constitutional AI; HH-RLHF; Responsible Scaling Policy) 3.2 OpenAI (GPT-Serie und o-Serie; RLHF; Reasoning-Training; Domain Experts Program) 3.3 Google DeepMind (Gemini; Constitutional Classifiers; Synthetic Data; SIMA / Reasoning) 3.4 Meta (Llama-Serie; Open-Weights-Strategie; Llama-Guard; Community Fine-Tuning) 3.5 xAI (Grok; X-Daten; Reasoning Mode) 3.6 Mistral (Open-Weights; europäischer Compliance-Fokus) 3.7 Aleph Alpha, Cohere, AI21, Reka, DeepSeek, Qwen — Kurzprofile 3.8 Vergleichsmatrix: Trainingsstrategie × Datenstrategie × Domain-Tiefe × Compliance × Frontier-Performance

Kapitel 4 — Wirkung des Human-AI-Trainings auf LLM-Qualität

4.1 Helpfulness, Honesty, Harmlessness — empirische Befunde 4.2 Faktentreue und Halluzinationsreduktion 4.3 Reasoning- und Werkzeugnutzung 4.4 Domain-Performance (Recht, Steuern, Finance, Medizin, Code) 4.5 Bias-, Toxizitäts- und Sicherheitsreduktion 4.6 Stil-, Formatierungs- und Tonalitätskontrolle 4.7 Persistente Schwächen und offene

Forschungsfragen (Sycophancy, Reward Hacking, Spec Gaming, Goodhart-Risiken) 4.8 Wirkungsbewertung in Drei-Szenarien-Bandbreiten (Konservativ / Base / Ambitioniert) für Performance-Lift je Domäne

| Kapitel 5 — Dienstleistungsmarkt „AI Training by Humans“

5.1 Marktdefinition und Abgrenzung (Data Labeling vs. Expert RLHF vs. Red Teaming vs. Synthetic Data Curation) 5.2 Wertschöpfungskette (Sourcing → Onboarding/Calibration → Production → QA → Delivery → Evaluation) 5.3 Anbietertypologie 5.3.1 Generalistische Labeling-Plattformen (Scale AI, Appen, Sama, iMerit, Labelbox, CloudFactory) 5.3.2 Expert-Crowd-Marktplätze (Mercor, Surge AI, Invisible, Outlier — Scale-Tochter, DataAnnotation.tech, Prolific) 5.3.3 Frontier-Lab-eigene Programme (Anthropic Expert Network, OpenAI Domain Experts, Google Quality Raters, Meta Llama Annotators) 5.3.4 Spezialisierte Boutiquen (Snorkel AI, Toloka, Centaur Labs, Karya, RWS TrainAI) 5.3.5 Red-Teaming-Spezialisten (Haize Labs, Gray Swan AI, Robust Intelligence — IBM) 5.3.6 Synthetic-Data-Anbieter (Datology AI, Datologyx, Gretel — NVIDIA, Mostly AI — AWS) 5.4 Tiefenprofile (je Anbieter ca. 1–2 Seiten): Mercor, Scale AI, Surge AI, Snorkel AI, Sama, Toloka, Outlier, Invisible, DataAnnotation, Prolific, Appen, Labelbox Inhaltsraster pro Anbieter: Gründung, HQ, Finanzierung/Bewertung, Mitarbeiter/Talentpool, Kundensegmente, Trainingsansatz, technologische Plattform, Compliance-Profil (SOC2/ISO/EU-AI-Act), Differenzierung, öffentlich genannte Kunden. 5.5 Vergütungsstrukturen 5.5.1 Stundensätze nach Domäne (z. B. STEM-Generalist 20–40 USD/h, Medizin/Recht/Quant Finance 50–200 USD/h, Frontier-Reasoning-Experten 75–300 USD/h) 5.5.2 Stückpreismodelle (per Task, per Pairwise, per Rubric) 5.5.3 Performance- und Calibration-Boni 5.5.4 Long-Term-Engagements (Retainer, „Trusted Trainer“-Programme) 5.5.5 Equity- und Revenue-Share-Modelle (selten, z. B. Mercor-Pilots) 5.6 Geschäftsmodelle der Anbieter 5.6.1 Marktplatz (Mercor, Surge, Prolific) 5.6.2 Managed Service (Scale AI, Sama, Appen) 5.6.3 Platform-as-a-Service (Labelbox, Snorkel, Toloka) 5.6.4 Vertikal-spezialisierte Boutiquen 5.6.5 Hybridmodelle (Plattform + Managed + Audit) 5.7 Auftraggeber 5.7.1 Frontier Labs (Anthropic, OpenAI, Google DeepMind, Meta, xAI, Mistral, Microsoft AI, Amazon AGI) 5.7.2 Hyperscaler und Enterprise (AWS, Azure, GCP, NVIDIA, Salesforce, IBM) 5.7.3 Domänenspezifische LLM-Bauer (Harvey, EvenUp, Hippocratic AI, Glean, Sierra, Cohere Enterprise) 5.7.4 Öffentliche Stellen (US AISI, UK AISI, EU AI Office als Evaluator-Auftraggeber) 5.8 Marktgröße und Wachstumsprojektion 5.8.1 Globaler Markt für AI-Trainingsdaten- und Annotation-Services (Drei-Szenarien-Bandbreite 2024 → 2030) 5.8.2 Sub-Segment „Expert-RLHF / Domain-Trainer“ — Volumen und Wachstumsrate 5.8.3 Regionale Anteile (USA, Indien, Philippinen, Kenia, EU, DACH) 5.8.4 Treiber und Hemmnisse (Frontier-CapEx, Reasoning-Skalierung, Synthetic-Data-Substitution, EU-AI-Act, Mindestlohn-Regulierung) 5.8.5 Sensitivitäten

Kapitel 6 — Fachgebietsspezifische Vorbereitung auf AI-Trainingstätigkeiten (modular)

Modul-Schema (verbindlich identisch für jedes Fachgebiet) — neue Fachgebiete werden als zusätzliches Unterkapitel 6.x mit folgender Struktur ergänzt:

6.x.1 Profil und Eignungsanforderungen (typische Qualifikationen, Berufserfahrung, Tooling, Sprachen) **6.x.2 Relevante Trainings-Tasks** (z. B. SFT-Aufsätze, Pairwise-Präferenzen, Rubrik-Scoring, Red-Teaming, Eval-Erstellung, Synthetic-Data-Validation) **6.x.3 Inhaltliche Vorbereitung** (Pflichtwissen, Standardwerke, Frameworks, regulatorischer Kontext) **6.x.4 Methodische Vorbereitung** (Prompt-Literacy, strukturiertes Schreiben, Begründungstiefe, Bias-Bewusstsein, Argumentationsketten, Goldstandard-Beispiele) **6.x.5 Plattform-Auswahl und Bewerbung** (welche Anbieter rekrutieren, Tests, Calibration, Onboarding) **6.x.6 Vergütungsbenchmark und Verhandlungsleitlinien** **6.x.7 Compliance und IP** (NDA, Datenschutz, EU-AI-Act, Berufsrecht — z. B. StBerG, WPO, BRAO; Interessenkonflikte) **6.x.8 Karriere-Optionen** (Side-Income, Vollzeit-Trainer, Senior Reviewer, Lead-Annotator, Eval-Architect, Spin-off in „Verticalized RLHF Boutique“)

Initiale Fachgebiete (Pflicht in v1.0): 6.1 Corporate Finance Advisory 6.2 Tax 6.3 Audit 6.4 Legal 6.5 Business Consulting 6.6 Computer Science

Reservierte Slots für künftige Erweiterungen (Platzhalter in HTML/DOCX/PPTX vorsehen, in v1.0 leer): 6.7 Strategy Consulting · 6.8 Investment Banking · 6.9 Compliance & Risk · 6.10 Medizin · 6.11 Pharma & Life Sciences · 6.12 Engineering · 6.13 Wirtschaftsprüfung Public Sector · 6.14 Geistes- und Sozialwissenschaften · 6.15 weitere.

Kapitel 7 — Strategische Implikationen

7.1 Für Frontier Labs (Make-or-Buy, vertikale Datenstrategien, In-House-Expert-Networks) 7.2 Für Beratungs- und Prüfungsgesellschaften (neues Geschäftsfeld „AI Training Services“ als margenstarkes Add-on; Risiken Talent-Abwanderung) 7.3 Für Einzelpersonen und Boutiquen (Side-Income, Spezialisierung, Reputationsaufbau, Spin-offs) 7.4 Für Bildungsanbieter, Kammern und Verbände (Curricula, Zertifizierungen, „AI-Trainer-Akademien“) 7.5 Regulatorische Implikationen (EU-AI-Act Annex III, Hochrisiko-Systeme, Auditpflichten; Arbeits- und Mindestlohnrecht in Schwellenländern)

| Kapitel 8 — Drei strategische Optionen je Stakeholder

Je Stakeholder (Frontier Lab / Beratungsgesellschaft / Einzelperson / Bildungsanbieter):
Option Konservativ — Option Base — Option Ambitioniert, jeweils mit Beschreibung,
Investitionsbedarf, Zeithorizont, erwartetem ROI/Nutzen und Risiken (analog Vorgehen in
Corporate Finance Advisory v2.0).

| Kapitel 9 — Methodik, KI-Einsatz, Limitationen, Reproduzierbarkeit

9.1 Methodik im Detail (Quellenklassen, Suchstrategie, Triangulation) 9.2 Einsatz von Claude
Opus 4.6 (Prompting, Iteration, menschliche Review-Schleife) 9.3 Limitationen (Verfügbarkeit
nicht-öffentlicher Vendor-Daten, Pricing-Volatilität) 9.4 Reproduzierbarkeit (Versionierte
Quellen, Stand der Recherche)

| Kapitel 10 — Kernquellen und Anhang

10.1 Quellen — zweiseitig wie im Designsystem 10.2 Glossar (RLHF, DPO, SFT, RLVR,
RLAIF, Constitutional AI, Synthetic Data, Calibration, Reward Hacking, Sycophancy, Eval,
Rubric) 10.3 Methodologie-Anhang (Reward-Model-Architekturen, Loss-Funktionen —
kompakt formal) 10.4 Anbieter-Vergleichstabellen (Long-Form)

4. Pflicht-Deliverables und Datei-Namensschema

Alle Deliverables in `public/human-ai-training/`. Datum 2026-05-12, Version v1-0.

| Tier | Format | Dateiname |
|--|-----------------|---|
| Themen-Indexseite | HTML | index.html |
| Ausführliche Studie | HTML | 2026-05-12_studie-vollfassung-v1-0.html |
| Ausführliche Studie | DOCX | 2026-05-12_studie-vollfassung-v1-0.docx |
| Ausführliche Studie | PDF | 2026-05-12_studie-vollfassung-v1-0.pdf |
| Ausführliche Studie (Markdown-Quelle) | MD | 2026-05-12_studie-vollfassung-v1-0.md |
| (Optional) Mittelfassung | MD/HTML/ PDF | 2026-05-12_studie-mittelfassung-v1-0.* |
| (Optional) Kompaktfassung | MD/HTML/ PDF | 2026-05-12_studie-kompaktfassung-v1-0.* |
| Executive Summary | HTML | 2026-05-12_executive-summary-v1-0.html |
| Executive Summary | DOCX | 2026-05-12_executive-summary-v1-0.docx |
| Executive Summary | PDF | 2026-05-12_executive-summary-v1-0.pdf |
| Executive Presentation | PPTX | 2026-05-12_executive-presentation-v1-0.pptx |
| Executive Presentation | PDF | 2026-05-12_executive-presentation-v1-0.pdf |
| Agent-Prompt (dieses Dokument) | MD/HTML/ PDF | 2026-05-12_agent-prompt-v1-0.* |

Zusätzlich: Ergänzung einer neuen `.branche-card` in `public/index.html` (Markup-Vorlage siehe Abschnitt 7).

5. Quantifizierungs- und Modellierungsvorgaben

1. **Drei-Szenarien-Logik.** Konservativ / Base / Ambitioniert für (a) globalen Markt für AI-Trainingsdienstleistungen 2024–2030, (b) Sub-Segment Expert-RLHF, (c) Performance-Lift je Domäne, (d) Vergütungsbandbreiten je Fachgebiet.
 2. **Annahmen-Memo.** Vor Schreibbeginn ein Annahmen-Memo (intern, im .md-Quelldokument als Appendix) mit Modellunternehmen, Kostenbandbreiten, Adoptionsszenarien, Datenstand.
 3. **Quantifizierungsmindertiefe.** Pro Quantifizierung: Treiber, Methode (Top-down/ Bottom-up), Sensitivitäten, Plausibilitätsanker (Vendor-Disclosures, Vergleichsmarkt).
 4. **Vergütungsbenchmark.** Pro Fachgebiet in Kapitel 6 eine Tabelle „Einstieg / Mittelfeld / Top-Tier“ in USD/h und (umgerechnet) EUR/h, mit Verweis auf Quelle.
 5. **Marktgrößen.** Mindestens drei unabhängige Quellen pro Marktgrößenangabe; Bandbreite statt Punktprognose.
 6. **Kein Punktprognose-Stil.** Alle Zukunftsaussagen als Bandbreite mit explizit benannten Annahmen.
-

6. Rechercheheuristiken und Quellenklassen

Pflicht-Quellenklassen (themenspezifisch):

Primärquellen — Branchenreports & Industrie: McKinsey „The state of AI“, BCG GenAI-Reports, Bain Tech Reports, Gartner Hype Cycle for AI / Magic Quadrant for Data Labeling, IDC Worldwide AI Spending Guide, Forrester Wave: Data Annotation Tools, Cognilytica / OpinionLab, Andreessen Horowitz „State of Generative AI“, Sequoia „Generative AI“, The Information, The Verge, Bloomberg, Reuters, Wired, MIT Technology Review.

Vendor-Disclosures und Modellkarten: Anthropic (Model Cards Claude 3/4-Familie, Responsible Scaling Policy, Constitutional AI Paper, RSP-Updates), OpenAI (GPT-4 / GPT-4o / o-Serie Model Cards, System Cards), Google DeepMind (Gemini Technical Reports, SIMA, AlphaProof), Meta (Llama 3 / 4 Papers), Mistral, xAI Grok Releases, DeepSeek-Papers, Qwen-Reports.

AI-Training-Anbieter: Mercor, Scale AI, Surge AI, Snorkel AI, Sama, Appen, Labelbox, Toloka, iMerit, Prolific, Invisible, DataAnnotation.tech, Outlier (Scale), CloudFactory, RWS TrainAI, Centaur Labs, Karya.

Akademische Quellen: arXiv (RLHF/DPO/RLAIF/Constitutional AI/RLVR), NeurIPS, ICML, ICLR, ACL, EMNLP, ICRL Safety-Workshops, Anthropic Research, OpenAI Research, DeepMind Research.

Regulierung, Recht & Statistik: EU AI Act (Verordnung 2024/1689) und Annex III, EU AI Office (Guidelines, Code of Practice), NIST AI RMF, US Executive Order on AI (2023; potenzielle Folge-Acts), UK AISI / US AISI Evaluations, OECD AI Policy Observatory, BSI/BfDI, BaFin (MaRisk AT 9 / BAIT / VAIT, Hinweise zu KI), IDW (PS/EPS/Hinweise zu KI in der Abschlussprüfung), Bitkom, eco, BVDS, Eurostat / Destatis, Bundesagentur für Arbeit (Stellenmarkt-Statistik).

Berufsrecht / Standesrecht: StBerG / BOSTB (Tax), WPO / BS WP/vBP (Audit), BRAO / FAO (Legal), BDU-Berufsgrundsätze (Consulting), ACM Code of Ethics (Computer Science).

7. Strukturvorgaben für `index.html` und `public/index.html`

| 7.1 `public/human-ai-training/index.html`

Pflichtabschnitte (1:1 nach Master `public/corporate-finance-advisory/index.html`): - `.site-header` mit Breadcrumb, `<h1>`, Untertitel, Logo, `.header-accent-bar` - **Prompt-Box** mit dem in Abschnitt 1 dieses Dokuments formulierten Original-Prompt (gekürzte Fassung) und Modell-Tag „Claude Opus 4.6“ - **Key Takeaways** (3–5 Bullets — siehe Abschnitt 8) - `.dl-bereich` mit drei `.content-tier`-Blöcken: 1. Executive Presentation (PPTX + PDF) 2. Executive Summary (HTML + DOCX + PDF) 3. Ausführliche Studie (HTML + DOCX + PDF; optional Mittel-/Kompaktfassung) - Zusätzlicher Tier „Methodik & Reproduzierbarkeit“ mit Link auf diesen Agent-Prompt - `.page-disclaimer` mit Standard-Disclaimer und Stand „Mai 2026“

| 7.2 Ergänzung in `public/index.html`

Neue `.branche-card` für `human-ai-training` einfügen (Beispiel-Schablone — Texte sind beim Erstellen final auszuarbeiten):

```
<a href="human-ai-training/index.html" class="branche-card">
  <h3>Human-AI-Training & Integration von Spezialwissen</h3>
  <p>Konzepte, Strategien der Frontier Labs, Dienstleistungsmarkt für „AI Training“</p>
  <div class="key-takeaways">
    <div class="key-takeaways-label">Key Takeaways</div>
    <ul>
```

```
<li>[Markt für AI-Trainingsdienstleistungen global – Bandbreite einsetzen]
<li>[Wachstumstreiber Reasoning-Training, Domain-Experts, Synthetic-Data-V
<li>[Fachgebietsspezifische Vorbereitung für CF, Tax, Audit, Legal, Consul
</ul>
</div>
<span class="badge">1 Studie · 1 Executive Summary · 1 Präsentation</span>
<div class="card-disclaimer">Kein Ersatz für Steuer-, Rechts- oder Anlageberat
</a>
```

8. Pflicht-Key-Takeaways (Platzhalter, im Erstellungsprozess final zu schärfen)

1. Globaler Markt für „AI Training by Humans“ wächst von [X] Mrd. USD (2024) auf [Y–Z] Mrd. USD (2030); Sub-Segment Expert-RLHF überproportional.
 2. Frontier Labs verlagern Trainingsaufwand zunehmend von generalistischem Labeling auf domänen- und reasoning-spezialisierte Experten; Mercor, Scale, Surge dominieren das Premium-Segment.
 3. Vergütungsbandbreiten für Domänen-Experten reichen von [a] USD/h (Generalist) bis [b] USD/h (Top-Tier Quant/Legal/Medizin); marktübliche Strukturen sind Marktplatz, Managed Service und Platform-as-a-Service.
 4. Corporate Finance Advisors, Steuerberater, Wirtschaftsprüfer, Anwälte, Unternehmensberater und Informatikerinnen haben fachgebietsspezifische, gut abgrenzbare Vorbereitungspfade — die Studie liefert je Fachgebiet ein Pflichtmodul nach einheitlichem Schema.
 5. Drei strategische Optionen je Stakeholder (Frontier Lab, Beratungsgesellschaft, Einzelperson, Bildungsanbieter) mit Konservativ-/Base-/Ambitioniert-Szenarien.
-

9. Executive Summary — inhaltliche Pflichtbestandteile

Layout 1:1 nach public/corporate-finance-advisory/2026-05-08_executive-summary-v2-0.html, mit Cover-Page (print-only) und Content-Section.

Pflichtabschnitte: 1. **Kernaussage** (1 Absatz, max. 120 Wörter) 2. **Markt in Zahlen** (Tabelle: globaler Markt 2024 / 2030 Base / 2030 Ambitioniert, Sub-Segment Expert-RLHF, Vergütungsbandbreiten je Fachgebiet, ROI-Bandbreiten je Stakeholder, Anteil USA / DACH / EU am globalen Volumen) 3. **Fünf zentrale Erkenntnisse** (je Erkenntnis 1 Absatz) 4.

Anbietervergleich Kurzübersicht (Tabelle: Mercor, Scale, Surge, Snorkel, Sama mit Spalten Modell · Spezialisierung · Vergütungsprofil · Kunden · Stärken) 5. **Fachgebietsspezifische Vorbereitung — Quick-Reference-Tabelle** (eine Zeile je Fachgebiet aus Kapitel 6, Spalten: Profil · Tasks · Vergütungsband · Top-Plattformen) 6. **Drei strategische Optionen pro Stakeholder** in Kurzform 7. **Methodik & KI-Einsatz** und **Kernquellen dieser Analyse** (Pflicht-Schluss-Sektionen) 8. **Disclaimer-Block** (Standard-Disclaimer + Vertraulichkeitsvermerk + Schlussvermerk)

10. Executive Presentation — Foliengerüst (Pflicht)

Layout 1:1 nach Designsystem Abschnitt 2 und Master public/corporate-finance-advisory/2026-05-08_executive-presentation-v2-0.pptx. 14–18 Folien.

1. Cover-Folie (Designsystem 2.1)
 2. Agenda
 3. Forschungsfragen und Methodik (Kurzfassung)
 4. Konzepte Human-AI-Training — Pipeline-Grafik
 5. Trainingsstrategien Frontier Labs — Vergleichsmatrix
 6. Wirkung des Human-AI-Trainings — Heatmap je Domäne
 7. Dienstleistungsmarkt — Wertschöpfungskette
 8. Anbieterlandschaft — Map / Quadrant
 9. Tiefenprofil Mercor (eine Folie)
 10. Vergütungsstrukturen und Geschäftsmodelle
 11. Auftraggeber-Mapping
 12. Marktgröße und Wachstumsbandbreite — Drei-Szenarien-Chart
 13. Fachgebietsspezifische Vorbereitung — Quick-Reference-Matrix
 14. Drei strategische Optionen je Stakeholder
 15. Methodik · KI-Einsatz · Rechtliche Hinweise (Pflicht-Schlussfolie)
 16. Kernquellen dieser Analyse (Pflicht-Schlussfolie)
-

11. Bilder, Grafiken und Tabellen

- **Logo** ausschließlich `public/assets/logo_katalysia_research.png` (Cover und Inhaltsfolien oben rechts).
 - **Grafiken** als Vektor-Grafik (SVG bzw. PPTX-Shapes); keine Stockfotos.
 - **Pflichtgrafiken:** (i) LLM-Trainings-Lebenszyklus (Kap. 2.1), (ii) RLHF-/DPO-Pipeline (Kap. 2.3), (iii) Wertschöpfungskette AI-Training-Services (Kap. 5.2), (iv) Anbieter-Quadrant (Kap. 5.3), (v) Marktgrößen-Bandbreite 2024–2030 (Kap. 5.8), (vi) Heatmap Performance-Lift je Domäne (Kap. 4.4), (vii) Fachgebiets-Matrix (Kap. 6).
 - **Tabellen** mit `.table-wrap`-Wrapper im HTML; einheitlich linksbündig, Bordeaux-Header.
-

12. Qualitätssicherung und Querprüfung

1. Vor Abschluss: Querprüfungs-Checkliste aus Designsystem Abschnitt 8 vollständig abarbeiten und im Übergabe-Memo dokumentieren.
 2. Inhaltliche Konsistenz: Zahlen, Aussagen, Annahmen, Modellversion und Datum identisch in PPTX, DOCX und HTML.
 3. Master-Querprüfung: Layout-Diff gegen `public/corporate-finance-advisory/` (Cover, Breadcrumb, Akzentlinien, Schlussfolien, HTML-Schluss-Sektionen, Branche-Card).
 4. Modularitätstest: Das Kapitel-6-Schema muss so geschrieben sein, dass ein neues Fachgebiet (z. B. „Strategy Consulting“) später als Copy-Paste-Block mit Anpassung der inhaltlichen Bullets eingefügt werden kann, ohne dass an Kapitel 1–5 oder 7–10 etwas geändert werden muss.
 5. Reproduzierbarkeit: Quellenliste, Stand der Recherche, Modellversion (Claude Opus 4.6) und Erstellungsdatum (2026-05-12) im Schlussvermerk.
-

13. Versionspflege regulatorischer Aussagen

Soweit die Studie Aussagen zu EU-AI-Act, Hochrisiko-Klassifizierung, NIST AI RMF, BAIT/VAIT/MaRisk, IDW-Hinweisen zu KI, DORA oder vergleichbarer Regulierung enthält, gilt verbindlich die Versionspflege-Routine gemäß `Versionspflege_EU-AI-Act_IDW.md`. Trigger, Watchlist und Versionsprotokoll sind im Schlussvermerk zu referenzieren.

14. Übergabenotiz (Pflicht beim Abschluss)

Im Übergabe-Memo (2026-05-12_uebergabe-notiz-v1-0.md im selben Themenordner) sind zu dokumentieren: - Inhaltsstand und Modellversion - Verwendete Hauptquellen je Kapitel - Annahmen und ihre Bandbreiten - Limitationen und nicht final geklärte Punkte - Vorschläge für künftige Fachgebiete in Kapitel 6 mit Priorisierung

15. Ablauf für den ausführenden Agenten (Quick-Reference)

1. Designsystem und Master-Referenzen lesen (KATALYSIA-Designsystem_Verbindlich.md, public/corporate-finance-advisory/).
 2. Annahmen-Memo erstellen.
 3. Vollfassung in Markdown drafte → DOCX → PDF.
 4. Executive Summary (HTML + DOCX + PDF) — inhaltlich aus Vollfassung kondensieren.
 5. Executive Presentation (PPTX + PDF) — Foliengerüst Abschnitt 10.
 6. Themen-Indexseite public/human-ai-training/index.html anlegen.
 7. public/index.html um neue .branche-card ergänzen.
 8. Querprüfungs-Checkliste abarbeiten.
 9. Übergabenotiz schreiben.
-

16. Standard-Textbausteine (verbindlich, gemäß Designsystem Abschnitt 5)

METHODIK & KI-EINSATZ (für diese Studie ausformuliert): 1. „Erstellt mit Claude Opus 4.6 auf Basis öffentlich zugänglicher Quellen (Branchenreports, Vendor-Dokumentation, Modellkarten, akademische Literatur, Anbieter-Webseiten, Stellenmarkt-Daten, Regulierungspublikationen).“ 2. „Schwerpunkt der Analyse: Globaler Markt für AI-Trainingsdienstleistungen, mit regionalen Vertiefungen USA, DACH und EU; Zeitraum 2022–2030, historische Vergleichsdaten ab 2018.“ 3. „Zahlenangaben zu Marktvolumina, Anbieter-Bewertungen, Mitarbeiterzahlen und Vergütungen basieren auf Anbieterangaben und Drittstudien und wurden nicht unabhängig verifiziert.“ 4. „Quantifizierungen sind szenarienbasiert (Konservativ / Base / Ambitioniert); tatsächliche Entwicklungen hängen von

Frontier-CapEx, Reasoning-Skalierung, Synthetic-Data-Substitution und Regulierungsdynamik ab." 5. „Stand der Recherche: Mai 2026 — spätere Marktbewegungen sind nicht berücksichtigt.“

DISCLAIMER, Vertraulichkeitsvermerk und Schlussvermerk: wörtlich gemäß Designsystem Abschnitt 5.2 / 5.3 / 5.4.

17. Erfolgskriterien (Definition of Done)

- Alle Pflicht-Deliverables in `public/human-ai-training/` vorhanden und in `public/index.html` verlinkt.
- Querprüfungs-Checkliste vollständig grün.
- Modularitätstest Kapitel 6 erfolgreich (neues Fachgebiet kann ohne Re-Architektur ergänzt werden).
- Inhaltliche Konsistenz Zahlen/Aussagen über alle Tier-Formate.
- Übergabenotiz dokumentiert Annahmen, Limitationen und Priorisierung künftiger Erweiterungen.
- Standards-Konformität mit `KATALYSIA-Designsystem_Verbindlich.md` und Versionspflege-Routine.

— Ende des Agent-Prompts v1.0 —